Digital Libraries and Institutional Repositories: Drivers for Organisational Change

Author/Contributor:
Amos, Howard B.

Event details:
International Conference on Digital Libraries (ICDL 2006)
New Delhi, India

Publication Date:
2006

DOI:
https://doi.org/10.26190/unsworks/612

License:
https://creativecommons.org/licenses/by-nc-nd/3.0/au/
Link to license to see what you are allowed to do with this resource.

Digital Libraries and Institutional Repositories: Drivers for Organisational Change

Howard B Amos. BA, Dip IM, AALIA, MACS

Director, Library Information Technology

University of New South Wales, Sydney Australia

Abstract

This paper reviews the digital library services that have developed over recent time and the role and function of the digital librarian. It considers these along with the development of institutional repositories and examines the opportunities for the development and expansion of services to the library community. The requirements of the repository and how these align with the open access scholarly communication objectives of early developers of repositories and how they fit within the hybrid library are also considered. The role the digital librarian plays in support of institutional repositories is examined to help identify the organisational change required by the library to support the services the digital librarian provides.

Introduction

Librarians invented concepts and established standards so that principles of organisation could be applied to stores of knowledge. The move from physical to digital stores has created new imperatives for our organisations. This paper considers the changes being implemented at the University of New South Wales (UNSW) Library to meet these challenges. To provide relevant online services to its community UNSW Library has established an institutional repository; maintains and provides access to digital collections in support of learning and teaching; manages digital higher degree theses; supports digital library research projects and provides hosting services for all of these. The Library also plays an active role in the work being undertaken to define the University's framework on e-Learning and e-Research; contributing and advising on matters of scalability, sustainability and accessibility. This paper outlines the development roadmap that UNSW Library is following as part of the organisational changes being instituted to meet the service delivery challenges of the next decade.

Digital Libraries and the Digital Librarian

Once we had Cataloguers and Reference Librarians. Today we also have the Digital Librarian who designs and builds services using digital resources to support the Library's primary objective: to provide access to authoritative and reliable knowledge stores. Digital librarians are involved in the activities of acquisition, description, enhancing discoverability and building access to information like other librarians. The knowledge stores they build need to be held in an environment that is stable, extensible and one that users find relevant and useable. This makes managing digital objects and seeing to their care different from the needs of physical objects. A different set of knowledge and skills is required with the advent of digital information and the Internet as information space. Librarians have had to develop specific technical skills, in areas such as database management, content management systems (managing digital assets and web content), web site design and digitisation, as well as an understanding of how these skills should be applied to best effect within the library service as a whole.

This combination of old and new skill sets saw libraries managing and organising digital collections using metadata to organise digitally held information. Librarians still catalogue, but now we also use metadata to "describe digital objects". We should note the differences between the two and consider why cataloguing as a description of activity doesn't transcend beyond describing resources held locally as this gives an insight into some of issues digital librarians must address. Campbell (Campbell 2006) notes that cataloguing processes don't scale well. The networking revolution of the last century meant we could share the load by distributing cataloguing effort across centralised bibliographic databases to minimise duplication. However, in today's digital world there are too many digital objects, too many types of objects and different concerns on how they are described for these processes to cope.

Digital objects require more complex asset management structures. As well as describing the information and intellectual content of the object to exploit digital objects to their full we need other information – information about the objects format, and transformation capabilities, information to preserve and extend its use and information on access and usage rights. The digital librarian has to deal with changing and emerging formats and the implications these will have on processes and workflows. Access and interface issues need to be addressed in an environment where digital rights management and licensing constraints play an important factor in managing access. New protocols such as the metadata encoding and transmission standard (METS) which provides a XML format for encoding different metadata necessary for the management of digital library objects address the complex and compound requirements needed to provide complete descriptions of digital objects. Within a METS document for example, there can be four types of metadata: technical metadata (information about an object's creation, format, and use), rights metadata (copyright and license information), descriptive and administrative metadata regarding any primary source from which the object derives, and digital provenance metadata (information regarding source/destination relationships between files, such as

master/derivative relationships and information regarding migrations/transformations). (Library of Congress 2005)

The digital librarian must consider and deploy initiatives such as METS so that they can describe and manage complex digital objects, build and support stable and extensible knowledge stores and carry out the acquisition, description, and access facilitation to provide information services to their community. When considering issues of rights to the use (and reuse) of materials concerns of data authenticity and data integrity also need to be addressed at the object level. Van de Sompel (Van de Sompel 2005) mentions numerous initiatives, including the Open Digital Rights Language (ODRL) initiative, MPEG21-REL and XRML and makes the point that machine-readable rights expressions are needed for the machine-to-machine interactions that will be deployed to extend the use of repositories.

Services at UNSW Library

(Kuny and Cleveland 1996) note that technological progress has changed how libraries do their work, but has not changed why libraries exist in the first place – connecting people with information.

Like all academic and research libraries UNSW Library has been exposed, and reacted to, changes in the scholarly communication cycle and how people use libraries to access information. Studies in the late 1990's (Voorjib 1999) on student and academic usage patterns show the majority believed searching the Internet or the Web provides sufficient information resources. Later work (Palmer 2006) shows how the Internet is influencing what information comes into play during the process of scholarly production. At the same time online services and increased connectivity are blurring the lines between "user" and "library" in the information space. This new shared information space is fostering the development of new resources for information access that assist researchers in identifying and finding sources of information.

In 2000 the University Librarian at UNSW circulated a paper entitled _Enhance the Learning Experience @ UNSW_. The document articulated the concept of the hybrid library and proposed that the Library 'provide students with seamless integrated access, independent of time or place, to digital and print information resources to enhance the learning experience and support the University's teaching activities.' (Bates 2000) Prior to the implementing the new Information and Resource Access Management System (IRAMS) the Library used a number of systems that had evolved over a considerable number of years. In the IRAMS RFP document, the Library sought systems that would support the management of the hybrid environment in which it now operates. Furthermore the Library recognised that there was a need to review and change the workflow processes it employs. It sought software that featured networking functions with support for modern interoperability and electronic data interchange protocols to improve these processes.

Implementation of software to support a system such as IRAMS requires more than the mechanical process of converting data and designing new interfaces. It needs a strong focus on change management processes that must have a coherent co-ordinated approach and address all aspects of the Library's activities. Three years after implementation an internal review carried out by the author on the information space UNSW Library now occupies points to the Library still being in the 1990's of "digital librarianship". The current UNSW Library space is still based on the pre-digital era, where access to resources is tightly integrated within the integrated library system platform. Service provision is one of multiple web front ends, each relatively standalone. We have deployed our library system as a "silo" application, with limited interconnections and interactions– the user comes to the Library catalogue and to Sirius, the name give to the access service for electronic and online resources. Purely digital service frameworks are only now beginning to coalesce and become clearer – RSS feeds, the use of blogs and social bookmarking sites are new nodes that are being explored to link users' personal workspaces and library resources. The objective now is to take our services and resources into the users' space.

As well as its integrated library system (housing the catalogue) and Sirius, UNSW Library supports a variety of services and environments to manage its digital assets and hosts these on a variety of platforms. In the early days of the hybrid library UNSW Library developed access services to reprints of University papers, including examination papers that were scanned and made accessible via a MySQL database. The Library leads the Australian and New Zealand initiative to develop and extend access to higher degree theses in digital format. The aim of the Australasian Digital Theses (ADT) programme, which now includes 31 Australian and New Zealand universities, was to establish a distributed database of digital versions of theses produced by postgraduate research students at our universities. The ideal behind the program is to provide access to, and promote, Australian research to the international community.   In December 2005, the databases included 5010 digitised theses. Some 15 Universities in Australasia, including UNSW, now mandate submission of theses in digital format and these are accessible from the Australasian Digital Theses site at http://adt.caul.edu.au

As contributing partners to the ARROW project (Australian Research Repositories Online to the World) UNSW Library is building repository services to serve the needs of the University. The ARROW project has been developing and testing software solutions to support best-practice institutional digital repositories. A wide range of digital content types will be managed in these repositories. This includes a potential path for the redevelopment of the ADT metadata repository. Content from the ARROW@UNSW repository will be incorporated into the National Library's ARROW discovery services (http://search.arrow.edu.au/apps/ArrowUI/) to extend discoverability.

ARROW is open-standards based and facilitates interoperability within and between participating institutions. (Payne 2005). "ARROW@UNSW" is using the results of this work to meet two separate business drivers – to provide information management facilities including discovery to delivery

services for UNSW research outputs and to support the research assessment process at UNSW. As an information management facility ARROW@UNSW will relieve academics and their faculty of the need to maintain an access and dissemination service (perhaps a web server on one of the faculty's systems.), as well as requiring them to manage their own content. An audit of 3 schools at UNSW in 2005 showed that there were a number of discovery/delivery problems, including circular references and broken links resulting in inaccessible documents. None of these resources were discoverable using protocols such as OAI-PMH (Open Archives Interface Protocol for Metadata Harvesting (www.openarchives.org))

UNSW Library is the lead in a project that is digitally aggregating thematic research material within the Visual Arts sphere. The Dictionary of Australian Artists Online (DAAO) draws from the work done by Prof. Joan Kerr on the Dictionary of Australian Artists (Kerr 1984, 1992). The online dictionary aims to reflect the entire landscape and history of artistic production in Australia. Stage one of the DAAO will contain over 5,000 biographies of Australian artists and will provide bibliographic data, contextual information, papers, images and audio. Relatively unexplored areas such as relationships between Indigenous and non-Indigenous art will be realised through the DAAO. (DAAO 2005)

Within the broader context of the Library's management of IT, there is the need to introduce structural change to incorporate the concept of technology as enabler to continue the IRAMS implementation work and to build coherent digital services that unify all the resources UNSW Library has at its disposal. There must be strong management and coordination of technology resources as different groups inter-operate and impact each other. It is incumbent on the Library's technology management function to coordinate and work through the inevitable conflicts in agendas, resource requirements and processes.

Digital Library Services and Institutional Repositories

The term "repository" has begun to be used in reference to some types of digital collections. Unlike a collection of digital objects housed in a traditional library database, institutional repositories are being used to capture original research and other intellectual property generated by an institution's constituent population (Crow 2002). Clifford Lynch (Lynch 2003) goes further. He sees institutional repositories as 'a set of services for the management and dissemination of digital materials created by the institution and its community members. It is most essentially an organisational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organisation and access or distribution.'

Identifying the elements of digital library services and comparing them to how institutional repositories have developed allows us to determine resourcing and organisational impacts. We need to ensure these endeavours can mesh to provide coherent digital library services to the University's community. This will also help define the changing role the digital librarian fulfils. There are various drivers behind

establishing repositories. Leaving aside the pan institutional subject based or thematic repositories, as well as enhancing access to resources and providing open access to content (including publication services and aiding scholarly communication) repositories are being used to hold content that represents the digital assets (including research outputs) of an institution. In this role they are a "digital asset management system" and provide the basis for curatorial and information management services for these objects on behalf of the institution. At UNSW the ARROW institutional repository has been established to hold any mix of objects that can be represented digitally. Using the FEDORA data store with a management layer called VITAL ARROW@ UNSW allows content to be deposited by the content owner, as well as by librarians, and is capable of managing compound objects. ARROW@UNSW is also capable of being harvested by any OAI-PMH complaint service.

The purist "Open Access" advocate of an institutional repository would argue that the primary (if not the only) objective of the repository is to maximise the impact of published research by making it as widely available as possible, fundamentally changing the scholarly communications model. In the context of the UNSW information space, ARROW@UNSW plays a major role in managing the research outputs of the University. As such it also contributes to the UNSW response to the Research Quality Framework (RQF), the Australian Government's initiative, to formulate a world's best practice framework for evaluating research quality and impact. Under the RQF institutions will nominate research groupings, which will provide evidence portfolios of outputs (the "best" four) from eligible researchers, and the full list of research outputs produced in the assessment period (currently slated to be six years). ARROW will provide the infrastructure and faculties that will enable assessment panels to access, review and provide a quality assessment of these outputs.

As well as developing new ways for librarians to describe digital resources, the curatorial requirements for these new knowledge stores have also changed. Library administrators now have to determine where the Library's curatorial responsibilities should begin and end. Should it continue to be the maintenance of the research paper, as it appeared in a peer reviewed journal? What about the data set that the research was based on, that now resides alongside the research paper in the institutional repository? Van de Sompel (Van de Sompel 2005) sees repositories growing expediential as their role in the scholarly communication value chain becomes more widely recognised. He sees this value chain as beginning with the registration of new knowledge from research outputs that are ingested into the repository. As this research is discoverable through such standards as the OAI-PMH harvesting protocol it is accessed and validated by different parties. This validation builds awareness and leads to new research outputs. The end of the chain occurs with the archiving/preservation of the knowledge. Libraries will become content nodes on the network, capturing the intellectual output and exposing it to the wider community.

The value of storing and managing the research dataset will become recognised as it is demonstrated that repositories, as both knowledge stores and data store, facilitate the creation of new knowledge by

allowing for non-anticipated use of research datasets. The digital librarian must therefore build skills in data management, access control policies as well as building and exposing metadata to ensure maximum discoverability. This should not be considered a conceptual leap from traditional librarian skills. After all, repositories are about facilitating the use of materials in many contexts – this is a core function of libraries that librarians have always provided. However, digital library services are not the same as establishing and managing an institutional repository. Digital library services involve the organisation and management of information beyond traditional physical publications to facilitate access to these digital objects. It may also include a process of digitising material or assembling digital objects into collections or mining data to build access to knowledge. Digital library services do not, in themselves, ensure that information is discoverable to users outside of the Library's immediate user base and its integrated library management system.

Digital Services Development Roadmap

As the digital information age matures and specialist technical skills become more widely available, what long term role emerges for the digital librarian? Dempsey (Dempsey 2006) identifies two themes identifiable in this new information space: One is the supply chain he describes as the discovery to delivery service framework – which looks at access services between a user and a distributed library resource encompassing resource discovery, linking to resources and requesting/receiving delivery. This is some way from the initial activities of digital librarians, who were concerned with digitising resources and making these available along side, or preferable via the integrated library system.

Dempsey writes that the development of the web and greater depth of connectivity has allowed libraries to move from a peripheral role to a central role in the information space that supports research, teaching and learning that Universities engender. Raymond Yee (Yee 2005) from the Interactive University Project, at the University of California, Berkeley has developed a tool called the Scholar's Box that encapsulates the concept. The Scholar's Box gives users "gather/create/share" functionality, enabling them to gather resources from multiple digital repositories in order to create personal and themed collections and other reusable materials that can be shared with others for teaching and research. This gathering and sharing brings together users with the resources and services used from the network.

Dempsey notes that the network has created a new dynamic of discovery and use around major hubs of information infrastructure: Google, Amazon iTunes and so on. They have aggregated supply (unified discovery and reduced transaction costs), aggregated demand (brought a large audience to bear), and are developing into platforms which help other applications reach their goals. Academic libraries must now provide relevant services in an environment where there is a cornucopia of content and information services. Full text, available now, at no direct cost, is becoming the expected norm. Our community members are changing the way they interact and engage in their research, teaching and learning activities. Most importantly, the amount of time our users are prepared to invest in learning

about and using our services is reducing and is now almost non-existent. The future of the academic library is challenged by these new dynamics. They have caused us to think about how to deliver and integrate services; the future of the academic library lies in how well it meshes with a whole range of related services. (Wainwright 2004)

We are moving to a networking environment (Dempsey calls it a "flatter network"), where the gap between the Web and business applications is narrowing. Applications occupy a smaller footprint within systems resources and working over the web is now just part of our business. Data is flowing more readily into and between user environments. Web services and RSS are important parts of connecting into user environments to provide on demand services, as are interactive tools and facilities such as wikis and social bookmarking services. Users no longer need to come to the Library, there has been a shift from needing to directly access the library catalogue and its other databases; through web site/portal entry to these services; and beyond to individual workflow and personal information space as the focus of interaction with resources. This is a natural consequence of moving more activities onto the web. UNSW library is still at the web site/portal stage. Planning has begun to move our services so they become more network orientated and progress to moving into the users' space.

Traditionally library users have had to adapt their workflow to the library. As the network becomes more important libraries need to adapt their services to the network workflows of the user. UNSW Library needs to develop services to cope with the fact that discovery and delivering services need to be instantaneous - attention spans are short and alternatives are more attractive. There are many demands on attention and many resources are available. Where attention is scarce, the Library needs to provide services which save time, which are capable of being used by members of our community to satisfy their personal information needs and fit into their workflows. Aggregating resources alone is not enough. Resources and services need to be tailored and moved into the users' environment in ways that support research, teaching and learning.

How do you tailor resources for particular courses or for particular technical environments? How do you make resources visible in search engine results? How do you provide links back from other discovery venues to the library, so that the user can actually get the resource of interest? How do you support metadata creation or document deposit in an extensible way? These are some of the questions we will have to answer as we define how we will deliver services into the users' information space. We need to be able to take our resources, combine them with others and present these services by allowing users to incorporate their own delivery method. We need to move beyond the situation where UNSW Library services are only seen by those members of the community that makes its way into the library web presence and manages to find what they are looking for .

Institutions charged with managing and providing access to large amounts of information (The California Digital Library (CDL), The US National Digital Library for Science and Education (NDSL) and the US Los Alamos National Laboratory (LANL) are three examples) have devoted resources to

examining the architectural requirements need to support services delivered in the users' information space. Smaller academic institutions such as the UNSW Library can draw from their experiences to model the building blocks required to build/support a coherent infrastructure that should then feed transparently into information services. Van de Sompel notes that many libraries will soon be faced with the need to create and provide services over different data stores, repositories and external resources. Repositories will expand rapidly as their role in the scholarly communication value chain becomes more widely recognised and they will need to facilitate the use of material in many contexts. Other resources (resources from publishers/aggregators, blogs and wikis) need to be connected as part of that value chain. The NSDL (Lagoze et al 2002) developed core architectural requirements for a digital library service: establishing their collections (data stores), building a metadata repository harvested from these; identifying search and discover and rights management services and user profiles to provide personalisation services. These interact using protocol and standards such as OAI, qualified Dublin Core and SDLIP from which base services could be identified and developed. The strategy of deploying foundation services from which you can build services is sound. We must all walk before we can run, however there is value in deploying as quickly as possible some elements or core services that can act as demonstrators and incubators and provide stimulus to the definition of the service development framework. This also has the advantage of identifying priority "core" services needed to satisfy users' primary needs, before expanding to meet emerging needs or to exploit new opportunities.

Focus must also be placed on building the more mundane skill sets needs to execute the new service delivery framework – expertise in developing service descriptions, specification and documentation writing as well as business analysis skills need to be nurtured. For an academic library the size of UNSW Library there are significant resource commitments to be made.

Conclusions drawn

Comparing digital library services with the business imperatives of institutional repositories allows us to determine resourcing and organisational impacts that, when addressed, will ensure these endeavours can mesh to provide coherent digital library services for the University's community. We need to build services that aggregate resources, provide straightforward and efficient access to these and delivery them into the users' own information space.

We need to build staff experience and skill sets to understand the workflow requirements and constraints of these new services. We will need to adopt, deploy and contribute to standards that will allow interoperability between components of our services. We need an understanding of standards and their relevance to the challenges we face and an appreciation of how they can assist in meeting our business needs. As the information environment becomes more complex and distributed we must move to new types of user-centric services and standards will be crucial to facilitating this.

UNSW Library has established work units charged with developing the required skills and with responsibility of meeting service delivery outcomes. A Digital Library Program Office has been created to develop projects and design services in conjunction with the Library IT Systems Infrastructure Group. Within the Library's Information Services Department a service delivery unit has been incorporated, responsible for identifying, assesing, specifying and deploying the new and emerging services that our community require. These two units will work closely together to review and assess our service framework; specify and design new services and facilities and then implementing them.

The staff in these units will define and design the service framework so that we can reduce the number of platforms we maintain, develop common services that can deploy across resources and into the users information space, linking disparate data stores using agreed metadata standards to describe objects and their usage. Standards based services will be developed to harvest metadata and build linking services in a scaleable and extensible manner and provide usage and rights management based on descriptions held at the object or remote resource level. To be relevant and useful, we will also need to build user profiling capabilities to personalise services. All these will need to be developed within web and network based services to take our library services into the users' space. These teams are now deploying as quickly as possible some services that can act as demonstrators and incubators and provide stimulus to the service development framework. This will enable us to identify "core" services needed to meet essential needs which can be prioritised, before expanding facilities and services to fulfil a complete service development framework.

References

Bates, Marion. 2000. **Enhancing the Learning Experience @ UNSW**. *University of New South Wales 2000.* Internal briefing paper.

Campbell, Jerry D. 2006 **Changing a Cultural Icon: The Academic Library as a Virtual Destination.** *Educause Review* Volume 41 (1)

Crow, R. 2002. **The case for institutional repositories: A SPARC position paper**
< http://www.arl.org/sparc/IR/ir.html >

Dempsey, Lorcan. 2006. **The (Digital) Library Environment: Ten Years After**. *Ariadne* Issue 46 January 2006 < http://www.ariadne.ac.uk/issue46/dempsey/ >

Dictionary of Australian Artists Online (DAAO) 2005. < http://www.daao.org.au/ > Accessed May 21 2006

Forsha, Harry I.1992. *The Pursuit of Quality through Personal Change.* Milwaukee. ASQC Quality Press. 191pp

Harnad, S. 1999. **Free at Last: The Future of Peer-Reviewed Journals.** *D-Lib Magazine December 1999* Volume 5 (12)

Heery, R. and Powell, A. 2006. **Digital Repositories Roadmap: looking forward.**
< http://www.jisc.ac.uk/uploaded_documents/rep-roadmap-v15.doc > Accessed May 26 2006

Kerr, Joan (ed) 1992. *The Dictionary of Australian artists : painters, sketchers, photographers and engravers to 1870.* Melbourne. Oxford University Press. 889 pp

Kuny, T. and Cleveland, G. **Digital Libraries: Myths and Challenges.** Paper delivered at the 62nd IFLA General Conference. Beijing China August 25-31, 1996. < http://www.ifla.org/IV/ifla62/62-kuny.pdf > Accessed April 4 2006

Lagoze, Carl et. al. 2002 **Core Services in the Architecture of the National Digital Library for Science Education (NSDL)** < http://arxiv.org/abs/cs.DL/0201025. > Accessed December 02 2005

Library of Congress 2005. **METS an Overview and Tutorial**
< http://www.loc.gov/standards/mets/METSOverview.v2.html > Access February 10 2006

Lynch, C. 2003. **Institutional repositories: Essential infrastructure for scholarship in the digital age**. *ARL Bimonthly Report* 226 February 2003

Palmer, Carole L. 2006. **Scholarly work and the shaping of digital access**. *Journal of the American Society for Information Science and Technology* Volume 56 (11)

Payne, Geoff 2005. **ARROW Institutional Repositories: A Report on the decisions and experiences of the ARROW Projects.** Paper delivered at the ALIA Information Online Conference, Sydney, 1 February 2005 < http://arrow.edu.au/docs/> Accessed on January 28 2006

Sreenivasulu, V. 2000. **The role of a digital librarian in the management of digital information systems (DIS).** *The Electronic Library* Volume 18 (1)

Van de Sompel, H. 2005. **Untitled I: challenges ahead**. *Presentation to the International Ex Libris Group, Olybris Seminar Series.* Kos Greece, April 18 2005

Voorjib, H. J. 1999 **Searching Scientific Information on the Internet: A Dutch Academic survey.** *Journal of the American Society for information Science.* Volume 50 (7)

Wainwright, E. 2004. **People, Networks, Books: New Strategies for University Academic Information and Service Delivery**. Paper delivered at the Biennial Conference of the Australian Library and Information Association 2004. Gold Coast Australia 21 – 24 September 2004 < http://conferences.alia.org.au/alia2004/pdfs/wainwright.e.paper.pdf > Accessed January 22 2006

Yee, Raymond 2005. Interactive University Project, University of California, Berkeley **ScholarsBox** < http://raymondyee.net/wiki/ScholarsBox > Accessed May 15 2006.